09/22/2020

MULTIVARIATE ANALYSIS ASSIGNMENT 2

TEAM INFORMATION

Ruixin Yang (RUID: 197000459)

Aishwarya Senthilvel (RUID: 199001269)

DESCRIPTION

The project we are going to analyze is about the effect of a marketing campaign of a commercial bank in Portugal. This campaign is about promoting the sales of the banks' term deposits. We want to know what influences client's decisions about buying term deposits. In this analysis, the factor we are looking at is what key factor captures client's decisions. And there are bunch of other factors affecting this decision from various respects.

Hypothesizes

Considering the problem, we hereby propose three hypothesizes that construct our analysis. Based on the previous hypothesizes, we revised them to make them more specific. We specified the relevant independent variables in each hypothesis first. Then we postulate the relationship between these independent variables and the dependent variable in parentheses. For some independent variable, their relationships are not clear since they are categorical.

Hypothesis 1: We assume that clients' characteristics is correlated to the decision of buying term deposits. These characteristics include clients' age (positive), job, marital status, loan default record (negative), ongoing housing loan (negative), ongoing personal loan (negative), and education level (positive). For the marital status, we assume that married people are more likely to purchase term deposits.

Hypothesis 2: We assume that clients' previous contacts are relevant to the decision of buying term deposits. The independent variables here include number of days that passed by after the client was last contacted from a previous campaign (negative), contact communication type, last contact month of year, last contact duration (negative), number of contacts both before and during the campaign (positive), and outcome of previous campaign on clients.

Hypothesis 3: We assume that the performance of markets is relevant to the decision of buying term deposits. The independent variables here involve employment variation rate (negative), consumer price index (CPI), consumer confidence index (negative), EUIRBOR 3-month rate (negative), and number of employees (positive). For CPI, we assume that its increasing rate is negatively correlated with the probability of buying term deposits.

09/22/2020

Key Performance Indicator

There are several points of measure to monitor performance. However, focusing on a few high-level metrics will help determine overall performance of the bank. As for the weight of each aspects, we propose that the significance of their coefficients may be a good perspective. If an indicator has a more statistically significant coefficient. It has larger effect on the dependent variable. As for the construction of the overall performance, we think that weighted average is sufficient to integrate indicators' effect.

KEY PERFOMANCE INDICATORS:

KEY RESULT AREA	KEY PERFORMANCE INDICATORS	DESCRIPTION
Lead Sources	Job	Determines capability of the lead to predict future revenue through the specific lead
Origination Value	Loan	Total Revenue earned for each loan over a given time
Lead	Campaign	Number of customers acquired by marketing influence
Lead Conversion	Poutcome	Finding the right customers, keeping them engaged and converting them into leads of the institution
Revenue	Cons_price_idx	Measure of the average change over time in prices paid by urban consumers for market basket of consumer goods and services
Overall Economy	Cons_conf_idx	Consumer Confidence Index is a good indicator od developments in the future reflecting current business conditions

Structural Pyramid Analysis Plan (SPAP)

1. S.M.A.R.T Goal:

We want to determine factors affecting clients' decisions of buying term deposits from a bank.

2. Measures of the dependent variable:

I. The binary variable that shows whether a client subscribes to a term deposit. (Data.world database, binary values summed by dates) Based on it, we calculate the ratio of 1 (yes to the term deposit).

II. The binary variable that shows whether a client subscribe a term deposit. (Data.world database, binary values summed by names of clients) Based on it, we calculate the ratio of 1 (yes to the term deposit).

III. The binary variable that shows whether a client subscribe a term deposit. (Data.world database, binary values summed by names of deposit products) Based on which, we calculate the ratio of 1 (yes to the term deposit).

3. Propose potential aspects for independent variables:

I. Do clients' characteristics specified above influence their decisions of buying term deposits?
II. Do clients' previous contacts information specified above influence their decisions of buying term deposits?
III. Does the performance of market specified above influence their decisions of buying term deposits?

4. Determine independent variables:

I. From hypothesis 1 and the first question above, we can tell that independent variables may contain age, type of job, marital status, default record, ongoing housing loan, ongoing personal loan, and education level.
II. From hypothesis 2 and the second question above, we can tell that independent variables may contain number of days that passed by after the client was last contacted from a previous campaign, contact communication type, last contact month of year, last contact duration, number of contacts both before and during the campaign, and outcome of previous campaign on clients.
III. From hypothesis 3 and the third question above, we can tell that independent variables may contain employment variation rate, consumer price index, consumer confidence index, EUIRBOR 3-month rate, and number of employees.

5. Evaluate the accessibility of the data:

I. We know the location: consumer price index, consumer confidence index, EUIRBOR 3-month rate, type of job, age, marital status, default record, ongoing housing loan, ongoing personal loan.
II. We only know the existence: number of employees, contact communication type, outcome of previous campaign on clients, number of days that passed by after the client was last contacted from the last campaign, number of contacts both before and during the campaign, last contact month of year,.
III. The existence is unclear: employment variation rate, last contact duration, education level.

6. For independent variables that the locations are known, here are the table of primitive analysis:

Since the y axis for every independent variable is the ratio of yes in purchasing the term deposits, we omit the column of y axis.

independent variable	Chart type	x-axis
consumer price index	line	consumer price index
consumer confidence index	line	consumer confidence index
EUIRBOR 3-month rate	line	EUIRBOR 3-month rate
last contact month of year	bar	last contact month of year
Number of days that passed by after the client was last contacted from a previous campaign	line	Number of days that passed by after the client was last contacted from a previous campaign
Number of contacts both before and during the campaign	bar	Number of contacts both before and during the campaign
age	bar	age
type of job	bar	type of job
marital status	bar	marital status
default record	bar	number of defaults
ongoing housing loan	bar	number of housing loans
ongoing personal loan	bar	number of personal loans

Data Dictionary

TABLE	COLUMN	DATA TYPE	REQUIRED	SENSITIVE	VALUES	DESCRIPTION
Bank	Age	Numeric				
Bank	Job	Categorical			Admin Blue-Collar Entrepreneur Housemaid Management Retired Self-Employed Services Student Technician Unemployed Unknown	Type of job
Bank	Marital	Categorical			Divorced Married Single Unknown	Divorced also covers widowed

Bank	Education	Categorical			Primary Secondary Tertiary	
Bank	Default	Categorical			No Yes Unknown	Has credit in default?
Bank	Balance					
Bank	Housing	Categorical			No Yes Unknown	Has housing loan?
Bank	Loan	Categorical			No Yes Unknown	Has personal loan?
Bank	Contact	Categorical			Cellular Telephone	Contact Communication Type
Bank	Day	Categorical			Mon Tue Wed Thu Fri	Last contact day of the week
Bank	Month	Categorical			Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec	Last contact month of year
Bank	Duration	Numeric				Last contact duration in seconds
Bank	Campaign	Numeric				Number of contacts performed during this campaign and for this client

Bank	Pdays	Numeric				Number of days that passed by after the client was last contacted from a previous campaign * 999 means client was not previously contacted
Bank	Previous	Numeric				Number of contacts performed before this campaign and for this client
Bank	Poutcome	Categorical			Failure Nonexistent Success	Outcome of the previous marketing campaign
Bank	y	Binary			Yes No	Has the client subscribed a term deposit
Bank Additional	Emp_var_rate					Quarterly indicator pf employment variation rate
Bank Additional	Cons_price_idx					Consumer price index - monthly indicator
Bank Additional	Cons_conf_idx					Consumer confidence index - monthly indicator
Bank Additional	euribor3m					Euribor 3 month rate - daily indicator
Bank Additional	Nr_employed					Number of employees - quarterly indicator